



Does spatial proximity drive norovirus transmission during outbreaks in hospitals?

| | |
|---------------------------------|---|
| Journal: | <i>BMJ Open</i> |
| Manuscript ID: | bmjopen-2013-003060 |
| Article Type: | Research |
| Date Submitted by the Author: | 16-Apr-2013 |
| Complete List of Authors: | Harris, John; Health Protection Agency, Gastrointestinal, Emerging and Zoonotic Diseases Lopman, Benjamin; Health Protection Agency, Gastrointestinal, Emerging and Zoonotic Diseases; Centers for Disease Surveillance and Control, Cooper, Ben; Mahidol University, Faculty of Tropical Medicine; University of Oxford, Nuffield Department of Clinical Medicine O'Brien, Sarah; University of Liverpool, Institute of Infection and Global Health |
| Primary Subject Heading: | Epidemiology |
| Secondary Subject Heading: | Infectious diseases |
| Keywords: | INFECTIOUS DISEASES, Epidemiology < INFECTIOUS DISEASES, Infection control < INFECTIOUS DISEASES |
| | |

SCHOLARONE™
Manuscripts

Title

Does spatial proximity drive norovirus transmission during outbreaks in hospitals?

Authors:

John P Harris * ^{1,4}, Ben S Cooper ^{2,3}, Ben A Lopman ¹, Sarah J O'Brien ⁴

1. Gastrointestinal Emerging and Zoonotic Diseases Department, Health Protection Agency, London, England

2. Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand and

3. Centre for Clinical Vaccinology and Tropical Medicine, Nuffield Department of Clinical Medicine, University of Oxford, Oxford, UK

4. University of Liverpool Institute of Infection and Global Health and National Consortium for Zoonosis Research.

* Corresponding Author address: Gastrointestinal Emerging and Zoonotic Diseases Department, Public Health England, 61 Colindale Avenue, Colindale, London NW9 5EQ, England. Email: john.harris@phe.gov.uk Tel +44 2083277403

Key words: Norovirus outbreak infection control spatial analysis

Word Count: 2845

Abstract

Objective.

To assess the role of spatial proximity, defined as patients sharing bays, in the spread of norovirus during outbreaks in hospitals.

Design

Enhanced surveillance of norovirus outbreaks between November 2009 and November 2011.

Methods.

Data were gathered during 149 outbreaks of norovirus in hospital wards from five hospitals in two major cities in England serving a population of 2 million. We used the time between the first two cases of each outbreak to estimate the serial interval for norovirus in this setting. This distribution and dates of illness onset were used to calculate epidemic trees for each outbreak. We then used a permutation test to assess whether proximity, for all outbreaks, was more extreme than would be expected by chance under the null hypothesis that proximity was not associated with transmission risk.

Results

Sixty-five outbreaks contained complete data on both onset dates and ward position. We estimated the serial interval to be 1.86 days (95% confidence intervals 1.6-2.2 days), and with this value found strong evidence to reject the null hypothesis that proximity was unimportant ($p < 0.001$). Sensitivity analysis using different values of the serial interval showed that there was evidence to reject the null hypothesis provided the assumed serial interval was less than 2.5 days.

Conclusion.

Our results provide evidence that patients occupying the same bay as patients with symptomatic norovirus infection are at an increased risk of becoming infected by these patients compared with patients elsewhere in the same ward.

Introduction.

Norovirus is the commonest cause of gastrointestinal infection worldwide [1]. There are between two and three million cases occurring each year in the United Kingdom [2,3]. Norovirus commonly presents as outbreaks of diarrhoea and vomiting and frequently reported in hospitals, care-homes, schools and cruise ships [4,5]. Outbreaks in hospitals are disruptive, often leading to ward or bay closures, staff sickness and cancelled operations [6]. The cost of nosocomial outbreaks of norovirus to the National Health Service (NHS) in England was estimated at £115 million in 2002/2003 [6]. More recently the cost in one region in Scotland was estimated at £1.2 million in the two norovirus seasons from 2007-2008 [7]. Understanding the benefits of infection control measures is problematic, because they are usually instigated as a package with several measures being implemented during an outbreak. While these interventions are based on sound infection control principles, evaluating their efficacy in trials is difficult and the published literature on norovirus outbreaks does not provide clear evidence of the effectiveness of infection control measures [5]. In observational studies early ward closure has been shown to shorten the mean duration of outbreaks [6, 7]. There is also evidence that vomiting and the resultant aerosols are important in transmitting the infection. People close to vomiting events, either by being close to the person who initially vomited, or by occupying the same area some time after the initial vomiting event, have a higher infection risk.[8-11] However, these analyses are based on single outbreaks or events that led to subsequent infections. Improved understanding of how norovirus spreads in closed environments could lead to better infection control procedures.

The aim of our study was to assess how spatial proximity to a norovirus case is associated with risk of acquiring norovirus gastroenteritis. Our hypothesis was that patients sharing bays (small self-contained areas within wards) with patients with symptomatic norovirus infection

were more likely to become infected compared to those who were in another bay or part of the affected ward.

Materials and Methods

Setting

We carried out enhanced surveillance of norovirus outbreaks from all in-patient wards in five tertiary care hospitals serving two cities in England, with a combined catchment of approximately two million people.

Surveillance data

We collected data during outbreaks from individual patients on date of onset of illness, symptoms (diarrhoea and/or vomiting), last date of illness for each patient, location on the ward at the time of the patients' symptoms onset (recorded as bed number and bay number) and also the ward type.

For two hospitals, information was recorded on specially designed forms that were completed by infection control staff and returned to the Health Protection Agency (HPA) from January 2008 to November 2011. Each month, we contacted the infection control lead at these hospitals asking about suspected or laboratory confirmed norovirus outbreaks and, if any had occurred, for the forms to be completed and returned. In three other hospitals infection control specialists had recorded these data items specified on outbreaks of norovirus during the season of 2007/2008. These data were entered into a study database during several visits to the hospitals. Data were collected from November 2007 to November 2011.

Patient location during outbreaks

We obtained ward plans for two of the five hospitals, which assisted in locating patients in the ward if only part of the information on patient location was recorded in the outbreak reports.

Definitions

Outbreaks were defined as two or more cases of diarrhoea and or vomiting of infectious origin on a ward occurring within two days of the first case suspected or confirmed to be due to norovirus.

A bay is a small self-contained area within a ward. Usually bays contain between two to eight beds. Bays are not the same as individual single bed occupied rooms. Proximity was defined as patients who share a bay.

Analytical framework

The analysis is based on a probabilistic reconstruction of chains of transmission (trees) based on the dates of illness onset for patients affected in outbreaks. It makes use of methods developed for SARS transmission and later applied to norovirus [12-15]. If we knew with certainty who acquired infection from whom it would be straightforward to quantify the role of proximity in norovirus outbreaks, for example, by using regression analysis. However, in practice, transmission events are unobserved, so instead we consider all possible infection trees consistent with the data. We used a previously described approach to calculate the probability, π_{ij} , that patient i was infected by patient j for each pair of infected patients in each outbreak based on onset times and the serial interval distribution (the serial interval is the time from onset of symptoms in case i to case j), without using proximity data [13,14]. We then used the matrix of π_{ij} values to simulate 1000 possible infection trees for each outbreak, by assigning the infector of patient i to be patient j with probability π_{ij} . In these simulations, we assumed that the case with the earliest onset time was the index case and had no infectors on the ward. If more than one patient had the earliest onset date in a given outbreak, we selected the index case from these patients with equal probability in each simulation. For each outbreak k , we used these 1000 simulations to produce a proximity metric, P_k , defined as

$$P_k = \sum_{i,j,l} \frac{s_{ijkl} p_{ijk}}{1000}$$

where s_{ijkl} is equal to 1 if patient i was infected by patient j in simulation l of outbreak k and is zero otherwise. The p_{ijk} terms measure proximity between patients i and j in outbreak k . In this application, we consider this to be a binary variable equal to 1 if patients i and j occupied the same bay at the time of first symptom onset of these patients. An overall proximity metric, P , is obtained by summing the P_k values.

If people in the same bay pose a greater risk of infecting each other this will tend to lead to larger values of the proximity metric (P). We compared this observed metric with values obtained if proximity was not associated with transmission. This distribution was derived by performing random permutations of the bays of the patients in each outbreak and calculating P_k as above for each outbreak. These values were again summed to give an overall proximity metric, S , when proximity was by assumption not an important factor. We repeated this for 1000 random permutations of the patient bays to obtain 1000 sampled proximity metrics. These 1000 S values therefore represent a sample from the distribution of proximity metrics that would be expected if proximity played no role in spreading the disease during the outbreak. By comparing the 1000 sampled values of S with the observed value P we can evaluate whether transmission is more (or less) likely to occur between patients in close proximity. If proximity is unimportant the observed value of P would be unlikely to be in the tails of the distribution of S . If proximity leads to increased transmission the observed P value would tend to be greater than most of the sampled S values. If proximity leads to decreased transmission (which could occur as a result of enhanced hygiene measures, for example) the observed P value is likely to be smaller than most of the sampled S values. Formally, we can perform a two-sided hypothesis test with a null hypothesis that proximity is not important

where the p-value is given by the proportion of sampled S values which are the more extreme than the value observed, P .

Serial intervals

A key input for constructing the transmission trees is the serial interval. Our best estimate came from the observed distribution of the difference between first and second onset dates for each outbreak. Often, more than one patient was ill on the first day of the outbreak, so we used the first date of illness onset in the next patient(s) for calculating the serial interval. This gave a mean serial interval of 1.86 days (median 1 day, 95% C.I. 1.6 - 2.2 days, obtained by bootstrapping).

We also performed sensitivity analyses using different assumptions. First, we used an estimate from a study of a community outbreak of norovirus at a scouting jamboree, giving a mean serial interval of 3.6 days, with an assumed gamma distribution [13]. We then considered gamma distributions with the same variance (4.1) but with the mean serial intervals varying between 0.5 and 5 days in half day increments. Data were analysed using R [16].

Results

Data were collected from 149 outbreaks in five hospitals between November 2007 and November 2011. These outbreaks affected 1,694 patients and 456 staff. The average duration of the outbreaks, determined as the first date of onset to the last date of onset, was 8.9 days (median 8; range 1-40 days). Outbreaks affected an average of 11.4 patients (median 11; range 1 – 30) and an average of 3.2 staff (median 2; range 0 – 20). Data from these outbreaks gave a mean serial interval of 1.86 days. Figure 1 shows the distribution of serial intervals from the observed data and the nearest fitting gamma distribution

The spatial modelling analysis used data from 65 outbreaks where all data (for both onset dates and position in ward when taken ill) were complete. The outbreak characteristics were

similar in these 65 outbreaks compared to the full dataset. The corresponding figures for these outbreaks were: average length of outbreak 9.5 days (median 8 days) average number of patients 11.9 (median 11) average number of affected staff 3.4 (median 2).

Proximity analysis

Figure 2 shows the observed proximity metric and the distribution of proximity metrics obtained under the assumption that proximity was not associated with transmission (from the simulated permutations). This shows how the proximity metrics observed relate to the distribution of proximity metrics if distance were not important. The dashed line indicates the observed proximity metric (P) and the bars indicate the distribution of proximity metrics from the simulated permutations. For the model using the serial interval taken from the observed onset dates, the observed metric is outside of the range of the simulated distributions and is highly statistically significant ($P = 153.34$, $p < 0.001$). With serial intervals of less than 2 days proximity is either outside or at the extreme right of the simulated proximity metrics and the p values ranged from < 0.001 for serial intervals of 0.5 days to 0.01 at a serial interval of 2 days. If we increase the assumed serial interval, the proximity metric moves to within the range expected from the simulated values, and at 3 days the p value was 0.2 (figure 2). Using the gamma probability distribution derived from a community outbreak by Heinje *et al* [13] (mean serial interval 3.06 days), the proportion of observed proximity values fell within the range that would be expected if proximity were not important.

The results show that the proximity metric (P) was larger than would be expected by chance under the null hypothesis (that proximity is not important) up to a serial interval of 2.5 days ($p=0.05$).

Discussion.

We have detected a strong association where patients who are in the same bay as patients who become ill have a higher probability of themselves becoming ill compared to patients in a different bay. The strength of this evidence is sensitive to the assumed serial interval distribution. We used values derived from the dates of onset of illness in patients during outbreaks on hospital wards. We also performed sensitivity analysis using a serial interval distribution derived from a study of norovirus in children [13]. However, because the degree to which this generalizes to a hospital setting is unclear (intuitively the high contact rates in hospitals would be expected to lead to shorter serial intervals) [17] we explored serial intervals from 0.5 days to four days, whilst constraining the variance. Our results show that for serial intervals of less than 2 days the observed effect of proximity (sharing a bay with someone else who was ill) is very significant ($p < 0.001$) and for serial intervals up to 2.5 days remained significant at the 5% level. This pattern was similar whether using the observed serial interval distribution from the outbreak data or using a parametric probability distribution.

Our study has some limitations. Although data collection were standardised it is often difficult to assess the accuracy of the date and place that patients were when they became ill. Specifically, accurate information on patients' positions on a ward was available for 47 percent of the outbreaks. The spatial analysis was undertaken on 65 outbreaks. In addition to the sensitivity analysis we also analysed the data by including outbreaks where onset dates of illness were complete but data on patient location were incomplete (where fewer than 10% of patient data on position was incomplete, 85 outbreaks). We dealt with missing values by allocating a completely separate bay for patients with missing data on location at time of onset. This approach is conservative in that it would underestimate the impact of proximity. Secondly, we removed the patients from the outbreaks if positional information was missing. Despite this limitation, the additional models indicated that that the results are robust to

different assumptions about missing data which is evidenced by slightly higher probabilities obtained when using records with complete information only (Table 1 and figure 3 appendix 1, supplementary information). As a check to demonstrate that the results were not an artefact of the statistical methods, we also ran the models on data where patient position was randomly assigned. This showed no pattern and the proximity measures were not significant for any of these models.

We used more than one approach to modelling the infection trees because of the lack of data on serial interval in norovirus outbreaks. Heinje *et al*'s method used data from child siblings at home. This was a useful starting point but is unlikely to be applicable to transmission in a hospital setting. Therefore, we derived gamma distributions for serial intervals from half a day to four days. The average incubation period for norovirus is considered to be between 24 hours to 48 hours [4, 18]. In our analysis the serial intervals of up to 2.5 days is likely to be a more appropriate time period than the analysis from Heinje *et al*.

Molecular analysis of stool samples could more definitively link outbreaks, which can help to reveal transmission networks [19, 20]. For example, in this study we have assumed that each ward outbreak was distinct, i.e. all cases within a ward were part of a chain of transmission, but this may not necessarily be true. It is possible for multiple introductions to occur, and some outbreaks may have spread from one ward to another. Genetic characterisation of samples from each ward during possible multiple outbreaks of norovirus would shed light on transmission events and lead to further insight about the direction of transmission, including the possibility that the virus can be moved around the hospital.

Our study focused on patients rather than staff. Our hypothesis was that symptomatic patients who vomit are most likely to contaminate the area close to them and other patients in their vicinity. Obtaining data on staff movements is much more complicated and would only really be practical in a detailed prospective study.

The importance of spatial proximity in propagating transmission is consistent with other recent studies[15, 21]. One study which used similar methods to calculate the infection trees [15] suggests that symptomatic individuals are likely to be the drivers of outbreaks of norovirus in hospital settings. Furthermore, the effective reproductive number was significantly higher for symptomatic patients compared to that for symptomatic staff. Norovirus transmission between people in close contact during sport, both within and between teams, has also been shown to occur [22] as well as airborne transmission through explosive vomiting [23]. One study demonstrated that successive staff working on an aircraft in which a member of the public had vomited also became sick [11]. Norovirus has a low infectious dose [4, 24,25] shedding virus occurs during episodes of vomiting, where the virus can become aerosolized and expose others in the vicinity. Therefore, closing the bay quickly, preventing movement to and from that bay, and immediately paying attention to cleaning areas nearby to initial vomiting events are likely to be effective in preventing further spread. The index of suspicion for patients who become ill should be high and implementing infection control interventions should not be delayed until the results of sampling are received, because this would increase morbidity and prolong the outbreak. New guidelines on controlling outbreaks of norovirus in hospitals and care homes recently released in the U.K. [26] move away from the need to close wards and operate on a “manage within bays” principle. Our study has shown that patients in close proximity to symptomatic patients are at increased risk of becoming infected by these patients.

Conclusions

We have shown a clear role of spatial proximity in the transmission of norovirus in hospital outbreaks. Increasing barriers to movement between bays by closing affected bays promptly would be effective in preventing further spread.

Acknowledgements

We would like to thank the infection control staff at the NHS hospitals for providing the data used in this study.

Funding statement

No financial support was received for this study. All authors report no conflicts of interest relevant to this article

Competing Interests

None

Data sharing

There is no unpublished data. The program for the statistical analysis is available in the supplementary material.

Contributorship

John Harris, Sarah O'Brien and Ben Lopman designed the study and analysis was conducted by John Harris, Ben Lopman and Ben Cooper. Ben Cooper and John Harris wrote the programming algorithm for statistical analysis. All authors were involved in interpretation of the data and drafting the article.

References

1. Patel MM, Hall AJ, Vinje J, et al. Noroviruses: a comprehensive review. *J Clin Virol* 2009;**44**:1-8

2. Phillips G, Tam CC, Conti S, et al. Community incidence of norovirus-associated infectious intestinal disease in England: improved estimates using viral load for norovirus diagnosis. *Am J Epidemiol*. 2010;**171**:1014-22

3. Tam CC, Rodrigues LC, Viviani L, et al. Longitudinal study of infectious intestinal disease in the UK (IID2 study): incidence in the community and presenting to general practice. *Gut* 2012;**61**:69-77

4. Estes MK, Prasad BV, Atmar RL. Noroviruses everywhere: has something changed? *Curr Opin Infect Dis*. 2006;**19**:467-74

5. Harris JP, Lopman BA, O'Brien SJ. Infection control measures for norovirus: a systematic review of outbreaks in semi-enclosed settings. *J Hosp Infect*. 2010;**74**:1-9

6. Lopman BA, Reacher MH, Vipond IB, et al. Epidemiology and cost of nosocomial gastroenteritis, Avon, England, 2002-2003. *Emerg Infect Dis*. 2004;**10**:1827-34

7. Danial J, Cepeda JA, Cameron F, et al, Templeton KE. Epidemiology and costs associated with norovirus outbreaks in NHS Lothian, Scotland 2007-2009. *J Hosp Infect*. 2011;**79**:354-8

8. Marks PJ, Vipond IB, Carlisle D, et al, Evidence for airborne transmission of Norwalk-like virus (NLV) in a hotel restaurant. *Epidemiol Infect*. 2000;**124**:481-7

9. Marks PJ, Vipond IB, Regan FM, et al, A school outbreak of Norwalk-like virus: evidence for airborne transmission. *Epidemiol Infect*. 2003;**131**:727-36

10. O'Neill PD, Marks PJ. Bayesian model choice and infection route modelling in an outbreak of Norovirus. *Stat.Med*. 2005;**24**:2011-24

11. Thornley CN, Emslie NA, Sprott TW, et al, Recurring norovirus transmission on an airplane. *Clin Infect Dis*. 2011;**53**:515-20

12. Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *Am J Epidemiol*. 2004;**160**:509-16
13. Heijne JC, Teunis P, Morroy G, et al. Enhanced hygiene measures and norovirus transmission during an outbreak. *Emerg Infect Dis*. 2009;**15**:24-30
14. Heijne JC, Rondy M, Verhoef L, et al. Quantifying transmission of norovirus during an outbreak. *Epidemiology* 2012;**23**:277-84
15. Sukhrie FH, Teunis P, Vennema H, et al. Nosocomial Transmission of Norovirus Is Mainly Caused by Symptomatic Cases. *Clin Infect Dis*. 2012;**54**:931-7.
16. R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
17. Kenah E, Lipsitch M, Robins JM. Generation interval contraction and epidemic data analysis. *Math Biosci*. 2008;**213**:71-9
18. Patel MM, Hall AJ, Vinje J, et al. Noroviruses: a comprehensive review. *J Clin Virol*. 2009;**44**:1-8
19. Allen DJ, Gray JJ, Gallimore CI, et al. Analysis of amino acid variation in the P2 domain of the GII-4 norovirus VP1 protein reveals putative variant-specific epitopes. *PLoS ONE* 2008;**3**:e1485 doi:10.1371/journal.pone.0001485
20. Sukhrie FH, Beersma MF, Wong A et al, Using molecular epidemiology to trace transmission of nosocomial norovirus infection. *J Clin Microbiol* 2011;**49**:602-6
21. Tuenis, P, Heijne, JCM, Sukhrie F, et al, Infectious disease transmission as a forensic problem: who infected whom? *J R Soc Interface* 2013;**6**:10 20120955. doi: 10.1098/rsif.2012.0955. published first online 6 February 2013.
22. Desai R, Yen C, Wikswo M, et al. Transmission of norovirus among NBA players and staff, winter 2010-2011. *Clin Infect Dis* 2011;**53**:1115-7

23. Sawyer LA, Murphy JJ, Kaplan JE, et al. 25- to 30-nm virus particle associated with a hospital outbreak of acute gastroenteritis with evidence for airborne transmission. *Am J Epidemiol* 1988;**127**:1261-71

24. Teunis PF, Moe CL, Liu P, et al. Norwalk virus: how infectious is it? *J Med Virol* 2008;**80**:1468-76

25. Atmar RL, Estes MK, The epidemiologic and clinical importance of norovirus infection. *Gastroenterol Clin North Am* 2006;**35**:275-90, viii

26. Guidelines for the management of norovirus outbreaks in acute and community health and social care settings. 2012

http://www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1317131639453 accessed 15/04/2013

Figure legends.

Figure 1. Serial interval distribution derived from onset dates of illness from observed outbreaks.

Figure 2. observed (dashed line) and distribution of expected (bars) proximity metrics for each serial interval.

For peer review only

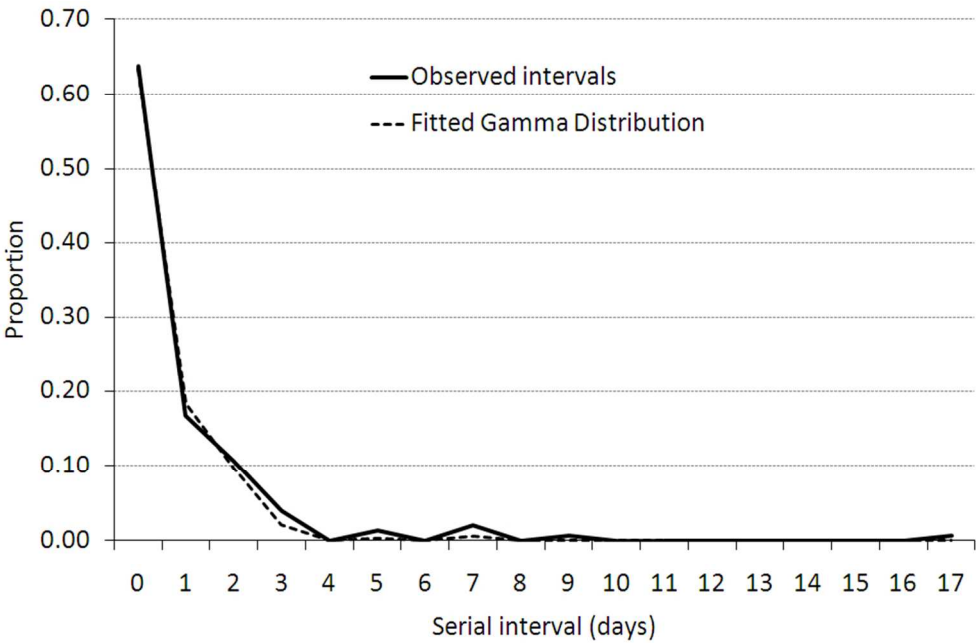


Figure 1. Serial interval distribution derived from onset dates of illness from observed outbreaks.

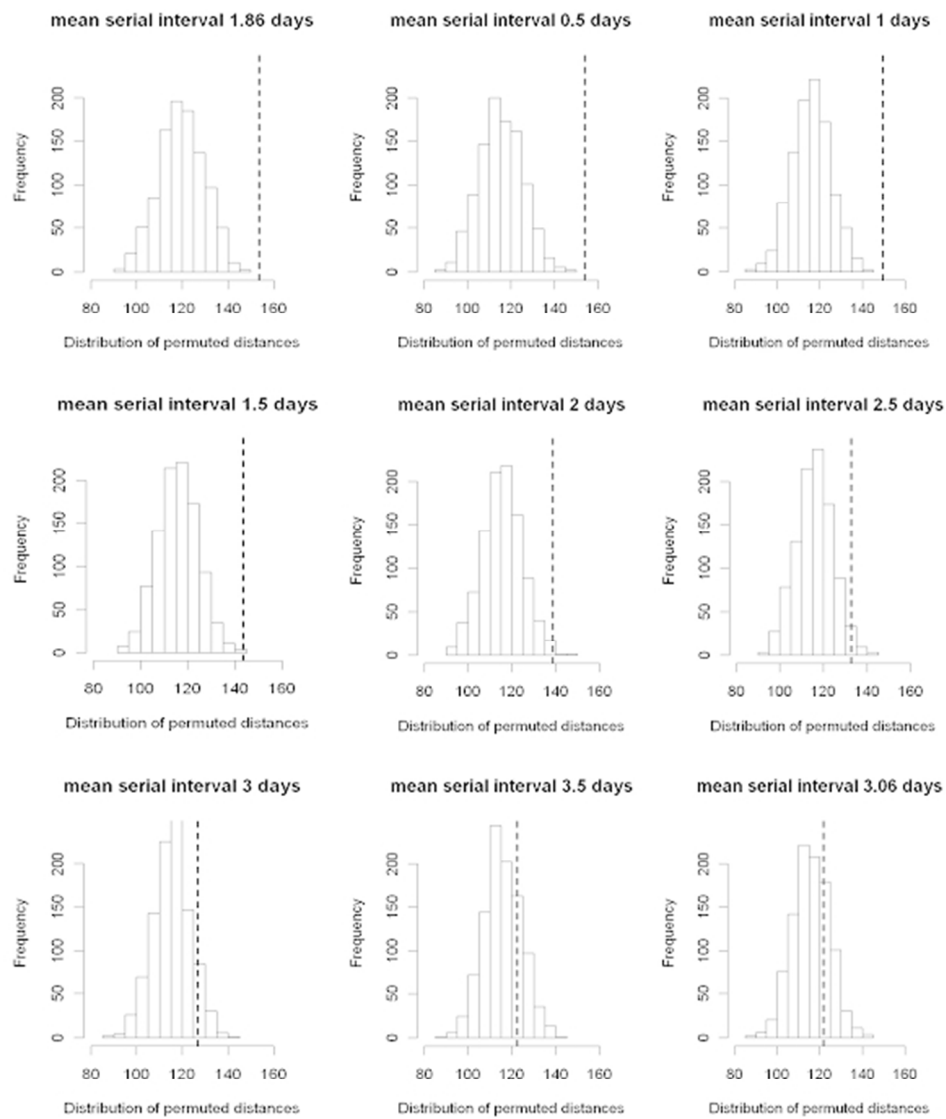


Figure 2. observed (dashed line) and distribution of expected (bars) proximity metrics for each serial interval.

Appendix 1 supplementary information

Table 1 Modelling approaches used in the analyses.

| | |
|--|---|
| Serial interval derived by: | |
| Method 1, from observed data on onset dates from hospital patients during outbreaks (N=149) | Method 2, using gamma distributions of serial intervals from Heinje et al (From healthy population) |
| Sensitivity analyses | |
| Analysis 1 (main analysis in results) Include only outbreaks with complete data on patient position at time of onset. (N=65) | |
| Analysis 2 Include outbreaks with missing values on patient position at time of onset. Patients allocated to distinct bays (N=85) | - |
| Analysis 3 Include outbreaks with missing values on patient position at time of onset, but exclude observations with missing values on patient position, i.e. patients deleted from dataset (N=85) | - |
| Analysis 4 Remove all information on patient position and randomly allocate bays | - |

Figure 1_supplement. Comparison of results from modelling approaches, main analysis and sensitivity analyses.

For peer review only

Appendix 2 supplementary material: R program code for calculations.

This first part of the program creates functions that will be called in the main part of the program later on.

create the distance matrix function which codes 1 for patients sharing a bay and 0 for those in different bays

```
makeDistancematrix<-function(pid,bay){
  n<-length(pid)
  Dist<-matrix(data=0, nrow=n,ncol=n)
  for(i in 1:n){
    for(j in 1:n){
      if(i!=j & bay[i]==bay[j]) Dist[i,j]<-1
    }
  }
  return(Dist)
}
```

Create the simulated outbreak function this simulates one outbreak

```
sim1outbreak<-function(pk1,onset.times.percase){
  n<-dim(pk1)[1]
  number.of.possible.index.cases<-sum(onset.times.percase==min(onset.times.percase))
  indexcase<-1+as.integer((number.of.possible.index.cases)*runif(1))
  whoinfectswhom<-rep(NA,n)
  for(i in 1:n){
    probnosource=max(1-sum(pk1[i,]),0)
    source<-which(rmultinom(1,1,c(pk1[i,],probnosource)) ==1)
    if(source==n+1) source<-0
    whoinfectswhom[i]<-source
    if(i==indexcase) whoinfectswhom[i]<-0 # i.e. infected outside
```

```
1
2
3 }
```

```
4
5 return(whoinfectswhom)
```

```
6
7 }
```

```
8
9 Create the function to randomise the bays in which patients are located
```

```
10
11 permutebays<-function(bays){
```

```
12
13 n<-length(bays)
```

```
14
15 return(bays[order(runif(n))])
```

```
16
17 }
```

```
18
19 Create the function to calculate the infection trees
```

```
20
21 calcOutbreakDist<-function(whoinfectswhom,Dist){
```

```
22
23 obDist=0
```

```
24
25 for(i in 1:length(whoinfectswhom)){
```

```
26
27 if(whoinfectswhom[i]!=0){
```

```
28
29 obDist=obDist+Dist[i,whoinfectswhom[i]]
```

```
30
31 }
```

```
32
33 }
```

```
34
35 return(obDist)
```

```
36
37 }
```

```
38
39 This section is the main program, it begins by reading in the data "data.csv" which is the comma
40 delimited data file containing the data.
```

```
41
42 dataset<-read.table("data.csv",header=T,sep=",")
```

```
43
44 dataset$dateonset<- as.Date(as.character(dataset$dateonset),"%d/%m/%Y")
```

```
45
46 S<-1000 [note this sets the number of simulated distance matrices to create]
```

```
47
48 numobks<-length(unique(dataset$obnumb))
```

```
49
50 realcumulativeDist<-rep(0,numobks) [to calculate cumulative distance over all simulated outbreak
```

```
51
52 trees]
```

```
53
54 permutedDistances<-matrix(data = rep(0,S*numobks), nrow = numobks, ncol = S)
```

```
55
56 for(obnumb in 1:numobks){
```

```
1
2
3 outbreak.id<-obnumb
4
5 onset.times.percase<- dataset$dateonset[which(dataset$obnumb==outbreak.id)]
6
7 unique.onset.times.percase<-unique(onset.times.percase)
8
9 unique.onset.times.percase<-1+unique.onset.times.percase - min(unique.onset.times.percase)
10
11 N<-length(onset.times.percase)
12
13 maxtimediff<-max(onset.times.percase)-min(onset.times.percase)
14
15 w<-rep(NA,maxtimediff+1)
16
17 w[1]<-0.6377
18
19 w[2]<-0.1678
20
21 w[3]<-0.1074
22
23 w[4]<-0.0403
24
25 w[5]<-0.0000
26
27 w[6]<-0.0134
28
29 w[7]<-0.0000
30
31 w[8]<-0.0201
32
33 w[9]<-0.0000
34
35 w[10]<-0.0067
36
37 w[11]<-0.0000
38
39 w[12]<-0.0000
40
41 w[13]<-0.0000
42
43 w[14]<-0.0000
44
45 w[15]<-0.0000
46
47 w[16]<-0.0000
48
49 w[17]<-0.0000
50
51 w[18]<-0.0067
52
53 w[19]<-0.0000
54
55 w[20]<-0.0000
56
57 w[21]<-0.0000
58
59
60
```



```

1
2
3 w[22]<-0.0000
4
5 w[23]<-0.0000
6
7 w[24]<-0.0000
8
9 w[25]<-0.0000
10
11 w[26]<-0.0000
12
13 w[27]<-0.0000
14
15 w[28]<-0.0000
16
17 w[29]<-0.0000
18
19 w[30]<-0.0000
20
21 w<-rep(NA,maxtimediff+1)
22
23 w[1]<-pgamma(0.5,shape=shape,scale=scale) # probability of a serial interval recorded as 0 days (i.e.
24 0 to .5 days)
25
26 for(i in 1:maxtimediff) w[i+1]<-pgamma(i+0.5,shape=shape,scale=scale)-sum(w[1:i])
27
28 wij<-matrix(rep(0,N^2),nrow=N)
29
30 number.of.possible.index.cases<-sum(onset.times.percase==min(onset.times.percase))
31
32 for(i in 1:N){
33
34   j<-1
35
36   onsettime<-onset.times.percase[j]
37
38   while(onsettime<=onset.times.percase[i] && j<=N){
39
40     timediff<-onset.times.percase[i]-onset.times.percase[j]
41
42     if(i!=j){
43
44       if(i> number.of.possible.index.cases){
45
46         wij[i,j]<- w[timediff+1]
47
48       } else {
49
50         wij[i,j]<- w[timediff+1]*(number.of.possible.index.cases-1)/ (number.of.possible.index.cases)
51
52       }
53
54     }
55
56   }
57
58   j<-j+1
59
60

```

```

1
2
3   onsettime<-onset.times.percase[j]
4
5   }
6
7   }
8
9   denoms<-c(NULL,rep(0,N-1)) # denoms (sum of likelihoods that person i was infected by person i-
10
11   1,i-2,...,1
12
13   for(i in 1:N){
14
15     denomfori<-0
16
17     for(j in 1:N) {
18
19       denomfori<-denomfori+wij[i,j]
20
21     }
22
23     denoms[i]<-denomfori
24
25   }
26
27   pkl<-matrix(rep(0,N^2),nrow=N) #relative likelihood that k was infected by l
28
29   for(l in 1:N){
30
31     for(k in 1:N){
32
33       numerator<-wij[k,l]
34
35       denom<-denoms[k]
36
37       if(denom>0){
38
39         pkl[k,l]<-numerator/denom
40
41       } else {
42
43         pkl[k,l]<-0
44
45       }
46
47     }
48
49   }
50
51   for(i in 1:number.of.possible.index.cases) pkl[i,<-pkl[i,]*(number.of.possible.index.cases-
52
53   1)/number.of.possible.index.cases
54
55   pid <- dataset$caseid[which(dataset$obnumb==outbreak.id)] #This limits the variable pid to one
56
57   outbreak
58
59
60

```

```

1
2
3 bay <- dataset$bay[which(dataset$obnumb==outbreak.id)] #This limits the variable bay to one
4
5 outbreak
6
7 Dist <- makeDistancematrix(pid,bay)
8
9 for (i in 1:S){
10
11 oneoutbreaksim<-sim1 outbreak(pk1,onset.times.percase) #a possible reconstruction of outbreak - who
12
13 infects whom.
14
15 realDist<-calcOutbreakDist(oneoutbreaksim, Dist)
16
17 realcumulativeDist[obnumb]<-realcumulativeDist[obnumb] + realDist
18
19 permD<-makeDistancematrix(pid,permutebays(bay)) #this is for the permuted bays
20
21 DistwithpermutedDistancematrix<-calcOutbreakDist(oneoutbreaksim, permD)
22
23 permutedDistances[obnumb,i]<-permutedDistances[obnumb,i]+ DistwithpermutedDistancematrix
24
25 }
26
27 realcumulativeDist[obnumb]<-realcumulativeDist[obnumb]/S
28
29 test.statistic<-realcumulativeDist[obnumb] # for this outbreak
30
31 pvalue.2sided<-
32
33 2*min(sum(permutedDistances[obnumb,]<=test.statistic)/length(permutedDistances[obnumb,]),sum(p
34
35 ermutedDistances[obnumb,]>=test.statistic)/length(permutedDistances[obnumb,]))
36
37 permutedDistancesAllOutbreaks<-colSums(permutedDistances)
38
39 test.statistic<-sum(realcumulativeDist) # for all outbreaks
40
41 pvalue.2sided<-
42
43 2*min(sum(permutedDistancesAllOutbreaks<=test.statistic)/length(permutedDistancesAllOutbreaks),
44
45 sum(permutedDistancesAllOutbreaks>=test.statistic)/length(permutedDistancesAllOutbreaks))
46
47 one.tailed.pvalue<-
48
49 sum(permutedDistancesAllOutbreaks>test.statistic)/length(permutedDistancesAllOutbreaks)
50
51 sink("outputfile.txt", append=FALSE, split=FALSE)
52
53 print(test.statistic)
54
55 print(pvalue.2sided)
56
57 sink()
58
59
60

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

```
jpeg("outputfile.jpeg", width=600, height=800, pointsize=30, quality=100)
hist(permutedDistancesAllOutbreaks,xlim=c(120,220), ylim=c(0,220), main="mean serial interval
3.06 days",
xlab = "Distribution of permuted distances")
abline(v=test.statistic, lty = 2, col="black", lwd = 3.5)
dev.off()
```

For peer review only

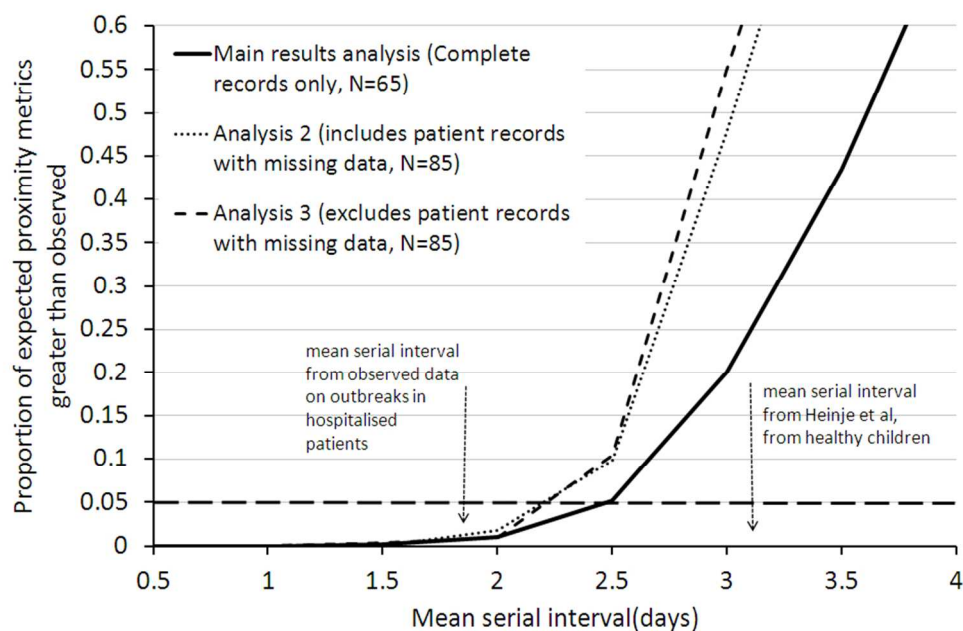


Figure 1_supplement. Comparison of results from modelling approaches, main analysis and sensitivity analyses

STROBE Statement—checklist of items that should be included in reports of observational studies

| | Item No | Recommendation |
|---------------------------|---------|--|
| Title and abstract | 1 | (a) Indicate the study’s design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found |
| Introduction | | |
| Background/rationale | 2 | Explain the scientific background and rationale for the investigation being reported |
| Objectives | 3 | State specific objectives, including any prespecified hypotheses |
| Methods | | |
| Study design | 4 | Present key elements of study design early in the paper |
| Setting | 5 | Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection |
| Participants | 6 | (a) Cohort study—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up Case-control study—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls Cross-sectional study—Give the eligibility criteria, and the sources and methods of selection of participants (b) Cohort study—For matched studies, give matching criteria and number of exposed and unexposed Case-control study—For matched studies, give matching criteria and the number of controls per case |
| Variables | 7 | Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable |
| Data sources/ measurement | 8* | For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group |
| Bias | 9 | Describe any efforts to address potential sources of bias |
| Study size | 10 | Explain how the study size was arrived at |
| Quantitative variables | 11 | Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why |
| Statistical methods | 12 | (a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) Cohort study—If applicable, explain how loss to follow-up was addressed Case-control study—If applicable, explain how matching of cases and controls was addressed Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses |

Continued on next page

Results

| | | |
|------------------|-----|---|
| Participants | 13* | (a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed (b) Give reasons for non-participation at each stage (c) Consider use of a flow diagram |
| Descriptive data | 14* | (a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders (b) Indicate number of participants with missing data for each variable of interest (c) <i>Cohort study</i> —Summarise follow-up time (eg, average and total amount) |
| Outcome data | 15* | <i>Cohort study</i> —Report numbers of outcome events or summary measures over time <i>Case-control study</i> —Report numbers in each exposure category, or summary measures of exposure <i>Cross-sectional study</i> —Report numbers of outcome events or summary measures |
| Main results | 16 | (a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period |
| Other analyses | 17 | Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses |

Discussion

| | | |
|------------------|----|--|
| Key results | 18 | Summarise key results with reference to study objectives |
| Limitations | 19 | Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias |
| Interpretation | 20 | Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence |
| Generalisability | 21 | Discuss the generalisability (external validity) of the study results |

Other information

| | | |
|---------|----|---|
| Funding | 22 | Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based |
|---------|----|---|

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at www.strobe-statement.org.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Does spatial proximity drive norovirus transmission during outbreaks in hospitals?

Article focus:

- Published literature on norovirus outbreaks does not provide clear evidence of the effectiveness of infection control measures
- Improved understanding of how norovirus spreads in closed environments could lead to better infection control procedures.
- This study uses statistical modelling methods to assess whether patients in close proximity are at increased risk of contracting norovirus during outbreaks in hospitals

Key Messages:

- A strong association where patients sharing a bay with patients who become ill have a higher probability of becoming ill compared to patients in a different bay
- Increasing barriers to movement between bays by closing affected bays promptly would be effective in preventing further spread

Strengths and Limitations of this study:

- Provides an estimation of serial interval, and assessment of significance of patient proximity in spreading norovirus within hospitals.
- Different modelling approaches showed consistent results.
- A weakness is that although data collection were standardised it is often difficult to assess the accuracy of the information on patients' positions on a ward.



Does spatial proximity drive norovirus transmission during outbreaks in hospitals?

| | |
|---------------------------------|---|
| Journal: | <i>BMJ Open</i> |
| Manuscript ID: | bmjopen-2013-003060.R1 |
| Article Type: | Research |
| Date Submitted by the Author: | 03-Jun-2013 |
| Complete List of Authors: | Harris, John; Health Protection Agency, Gastrointestinal, Emerging and Zoonotic Diseases Lopman, Benjamin; Health Protection Agency, Gastrointestinal, Emerging and Zoonotic Diseases; Centers for Disease Surveillance and Control, Cooper, Ben; Mahidol University, Faculty of Tropical Medicine; University of Oxford, Nuffield Department of Clinical Medicine O'Brien, Sarah; University of Liverpool, Institute of Infection and Global Health |
| Primary Subject Heading: | Epidemiology |
| Secondary Subject Heading: | Infectious diseases, Gastroenterology and hepatology |
| Keywords: | INFECTIOUS DISEASES, Epidemiology < INFECTIOUS DISEASES, Infection control < INFECTIOUS DISEASES |
| | |

SCHOLARONE™
Manuscripts

Title

Does spatial proximity drive norovirus transmission during outbreaks in hospitals?

Authors:

John P Harris * ^{1,4}, Ben A Lopman ¹, Ben S Cooper ^{2,3}, Sarah J O'Brien ⁴

1. Gastrointestinal Emerging and Zoonotic Diseases Department, Health Protection Agency, London, England

2. Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand and

3. Centre for Clinical Vaccinology and Tropical Medicine, Nuffield Department of Clinical Medicine, University of Oxford, Oxford, UK

4. University of Liverpool Institute of Infection and Global Health and National Consortium for Zoonosis Research.

* Corresponding Author address: Gastrointestinal Emerging and Zoonotic Diseases Department, Public Health England, 61 Colindale Avenue, Colindale, London NW9 5EQ, England. Email: john.harris@phe.gov.uk Tel +44 2083277403

Key words: Norovirus outbreak infection control spatial analysis

Word Count: 3236

Abstract

Objective.

To assess the role of spatial proximity, defined as patients sharing bays, in the spread of norovirus during outbreaks in hospitals.

Design

Enhanced surveillance of norovirus outbreaks between November 2009 and November 2011.

Methods.

Data were gathered during 149 outbreaks of norovirus in hospital wards from five hospitals in two major cities in England serving a population of 2 million. We used the time between the first two cases of each outbreak to estimate the serial interval for norovirus in this setting. This distribution and dates of illness onset were used to calculate epidemic trees for each outbreak. We then used a permutation test to assess whether proximity, for all outbreaks, was more extreme than would be expected by chance under the null hypothesis that proximity was not associated with transmission risk.

Results

Sixty-five outbreaks contained complete data on both onset dates and ward position. We estimated the serial interval to be 1.86 days (95% confidence intervals 1.6-2.2 days), and with this value found strong evidence to reject the null hypothesis that proximity was not significant ($p < 0.001$). Sensitivity analysis using different values of the serial interval showed that there was evidence to reject the null hypothesis provided the assumed serial interval was less than 2.5 days.

Conclusion.

Our results provide evidence that patients occupying the same bay as patients with symptomatic norovirus infection are at an increased risk of becoming infected by these patients compared with patients elsewhere in the same ward.

Introduction.

Norovirus is the commonest cause of gastrointestinal infection worldwide [1]. There are between two and three million cases occurring each year in the United Kingdom [2,3]. Norovirus commonly presents as outbreaks of diarrhoea and vomiting and are frequently reported in hospitals, care-homes, schools and cruise ships [4,5]. Outbreaks in hospitals are disruptive, often leading to ward or bay closures, staff sickness and cancelled operations [6]. The cost of nosocomial outbreaks of norovirus to the National Health Service (NHS) in England was estimated at £115 million in 2002/2003 [6]. More recently the cost in one region in Scotland was estimated at £1.2 million in the two norovirus seasons from 2007-2008 [7]. Understanding the benefits of infection control measures is challenging, because they are usually instigated as a package with several measures being implemented during an outbreak. While these interventions are based on sound infection control principles, evaluating their efficacy in trials is difficult and the published literature on norovirus outbreaks does not provide clear evidence of the effectiveness of infection control measures [5]. In observational studies early ward closure has been shown to shorten the mean duration of outbreaks [6, 7]. There is also evidence that vomiting and the resultant aerosols are important in transmitting the infection. People exposed to vomiting events, either by being close to the person who initially vomited, or by occupying the same area sometime after the initial event, have a higher infection risk.[8-11] However, these analyses are based on single outbreaks or events that led to subsequent disease. Improved understanding of how norovirus spreads in closed environments could lead to better infection control procedures.

The aim of our study was to assess how spatial proximity to a norovirus case is associated with risk of acquiring symptomatic norovirus gastroenteritis. Our hypothesis was that patients sharing bays (small self-contained areas within wards) with patients with symptomatic

norovirus infection were more likely to become infected compared to those who were in another bay or part of the affected ward.

Materials and Methods

Setting

We carried out enhanced surveillance of norovirus outbreaks from all in-patient wards in five tertiary care hospitals serving two cities in England, with a combined catchment of approximately two million people.

Surveillance data

We collected data during outbreaks from individual patients on date of onset of illness, symptoms (diarrhoea and/or vomiting), last date of illness for each patient, location on the ward at the time of the patients' symptoms onset (recorded as bed number and bay number) and also the ward type.

For two hospitals, information was recorded on specially designed forms that were completed by infection control staff and returned to the Health Protection Agency (HPA) from January 2008 to November 2011. Each month, we contacted the infection control lead at these hospitals asking about suspected or laboratory confirmed norovirus outbreaks and, if any had occurred, and for the forms to be completed and returned. In three other hospitals the data were downloaded from a database on which infection control specialists had recorded these data items during outbreaks of norovirus occurring in the season of 2007/2008. Data from these three hospitals were downloaded during several visits to these hospitals. Data on outbreaks on norovirus were available from November 2007 to November 2011.

Patient location during outbreaks

We obtained ward plans for two of the five hospitals, which assisted in locating patients in the ward if only part of the information on patient location was recorded in the outbreak reports.

Definitions

Outbreaks were defined as two or more cases of diarrhoea and or vomiting of infectious origin on a ward occurring within two days of the first case suspected or confirmed to be due to norovirus. The hospitals in this study all used Polymerase Chain Reaction (PCR) for detection of norovirus in stool samples.

A bay is a small self-contained area within a ward. Usually bays contain between two to eight beds. Bays are not the same as individual single bed occupied rooms. Proximity was defined as patients who share a bay.

Analytical framework

The analysis is based on a probabilistic reconstruction of chains of transmission (trees) based on the dates of illness onset for patients affected in outbreaks. It makes use of methods developed for SARS transmission and later applied to norovirus [12-15]. If we knew with certainty who acquired infection from whom it would be straightforward to quantify the role of proximity in norovirus outbreaks, for example, by using regression analysis. However, in practice, transmission events are unobserved, so instead we consider all possible infection trees consistent with the data. We used a previously described approach to calculate the probability, π_{ij} , that patient i was infected by patient j for each pair of infected patients in each outbreak based on onset times and the serial interval distribution (the serial interval is the time from onset of symptoms in case i to case j), without using proximity data. The serial interval distribution tells us the probability of durations of 0, 1, 2, ... days between onset in a case and onset in secondary cases infected by this case. Given multiple possible sources for a case, we can use knowledge of this distribution to tell us how likely each is to be the true source. Full technical details are described in Wallinga & Teunis (2004) [12].

We then used the matrix of π_{ij} values to simulate 1000 possible infection trees for each outbreak, by assigning the infector of patient i to be patient j with probability π_{ij} .

In these simulations, we assumed that the case with the earliest onset time was the index case and had no infectors on the ward. If more than one patient had the earliest onset date in a given outbreak, we selected the index case from these patients with equal probability in each simulation. For each outbreak k , we used these 1000 simulations to produce a proximity metric, P_k , defined as

$$P_k = \sum_{i,j,l} \frac{s_{ijkl} p_{ijk}}{1000}$$

where s_{ijkl} is equal to 1 if patient i was infected by patient j in simulation l of outbreak k and is zero otherwise. The p_{ijk} terms measure proximity between patients i and j in outbreak k . In this application, we consider this to be a binary variable equal to 1 if patients i and j occupied the same bay at the time of first symptom onset of these patients. An overall proximity metric, P , is obtained by summing the P_k values. The value of P (and of P_k for individual outbreaks) should be interpreted as a measure of how much transmission occurs between patients in the same bay.

If people in the same bay pose a greater risk of infecting each other this will tend to lead to larger values of the proximity metrics, P , and P_k . We compared this observed metric with values obtained if proximity was not associated with transmission. This distribution was derived by performing random permutations of the bays of the patients in each outbreak and calculating P_k as above for each outbreak. These values were again summed to give an overall proximity metric, S , when proximity was by assumption not an important factor. We repeated this for 1000 random permutations of the patient bays to obtain 1000 sampled proximity metrics. These 1000 S values therefore represent a sample from the distribution of proximity metrics that would be expected if proximity played no role in spreading the disease during the outbreak. By comparing the 1000 sampled values of S with the observed value P we can evaluate whether transmission is more (or less) likely to occur between patients in

close proximity. If proximity is unimportant the observed value of P would be unlikely to be in the tails of the distribution of S . If proximity leads to increased transmission the observed P value would tend to be greater than most of the sampled S values. If proximity leads to decreased transmission (which could occur as a result of enhanced hygiene measures, for example) the observed P value is likely to be smaller than most of the sampled S values. Formally, we can perform a two-sided hypothesis test with a null hypothesis that proximity is not important where the p-value is given by the proportion of sampled S values which are the more extreme than the value observed, P .

Serial intervals

A key input for constructing the transmission trees is the serial interval. Our best estimate came from the observed distribution of the difference between first and second onset dates for each outbreak, and our primary analysis made use of this empirical serial interval distribution. Often, more than one patient was ill on the first day of the outbreak, so we used the first date of illness onset in the next patient(s) for calculating the serial interval. This gave a mean serial interval of 1.86 days (median 1 day, 95% C.I. 1.6 - 2.2 days, obtained by bootstrapping).

We also performed sensitivity analyses using different assumptions. First, we used an estimate from a study of a community outbreak of norovirus at a scouting jamboree, giving a mean serial interval of 3.6 days, with an assumed gamma distribution [13]. We then considered gamma distributions with the same variance (4.1) but with the mean serial intervals varying between 0.5 and 5 days in half day increments. Data were analysed using R [16].

Results

Data were collected from 149 outbreaks in five hospitals between November 2007 and November 2011. These outbreaks affected 1,694 patients and 456 staff. The average duration

of the outbreaks, determined as the first date of onset to the last date of onset, was 8.9 days (median 8; range 1-40 days). Outbreaks affected an average of 11.4 patients (median 11; range 1 – 30) and an average of 3.2 staff (median 2; range 0 – 20). Data from these outbreaks gave a mean serial interval of 1.86 days. Figure 1 shows the distribution of serial intervals from the observed data and the nearest fitting gamma distribution

The spatial modelling analysis used data from 65 outbreaks where all data (for both onset dates and position in ward when taken ill) were complete. The outbreak characteristics were similar in these 65 outbreaks compared to the full dataset. The corresponding figures for these outbreaks were: average length of outbreak 9.5 days (median 8 days) average number of patients 11.9 (median 11) average number of affected staff 3.4 (median 2). The outbreaks affected various ward types, with most occurring in general medical wards (34%) and care of the elderly wards (28%). Other specialties were respiratory medicine (12%), stroke/neurology wards (11%), coronary care wards (9%), and orthopaedic/trauma wards (6%).

Proximity analysis

Figure 2 shows the observed proximity metric and the distribution of proximity metrics obtained under the assumption that proximity was not associated with transmission (from the simulated permutations). This shows how the proximity metrics observed relate to the distribution of proximity metrics if distance were not important. The dashed line indicates the observed proximity metric (P) and the bars indicate the distribution of proximity metrics from the simulated permutations. For the model using the serial interval taken from the observed onset dates, the observed metric is outside of the range of the simulated distributions and is highly statistically significant ($P = 153.34$, $p < 0.001$). With serial intervals of less than 2 days proximity is either outside or at the extreme right of the simulated proximity metrics and the p values ranged from < 0.001 for serial intervals of 0.5 days to 0.01 at a serial interval of 2 days. If we increase the assumed serial interval, the

proximity metric moves to within the range expected from the simulated values, and at 3 days the p value was 0.2 (figure 2). Using the gamma probability distribution derived from a community outbreak by Heijne *et al* [13] (mean serial interval 3.06 days), the proportion of observed proximity values fell within the range that would be expected if proximity were not important.

The results show that the proximity metric (*P*) was larger than would be expected by chance under the null hypothesis (that proximity is not important) up to a serial interval of 2.5 days ($p=0.05$).

Discussion.

We have detected a strong association where patients who are in the same bay as patients who become ill have a higher probability of themselves becoming ill compared to patients in a different bay. In other words, transmission of norovirus infections is more likely to occur among patients sharing a bay, compared to transmission among patients in different bays. Whilst this might at first seem an obvious finding, there are competing theories about the transmission of the virus in complex healthcare settings. For example transmission might occur through staff transferring virus on their hands or patients touching infected surfaces with their hands when moving around the wards or the hospital. The strength of our conclusion is sensitive to the assumed serial interval distribution. We used values derived from the dates of onset of illness in patients during outbreaks on hospital wards. We also performed sensitivity analysis using a serial interval distribution derived from a study of norovirus in children [13]. However, because the degree to which this generalizes to a hospital setting is unclear (intuitively the high contact rates in hospitals would be expected to lead to shorter serial intervals) [17] we explored serial intervals from 0.5 days to four days, whilst constraining the variance. Our results show that for serial intervals of less than 2 days the observed effect of proximity (sharing a bay with someone else who was ill) is highly

significant ($p < 0.001$) and for serial intervals up to 2.5 days remained significant at the 5% level. This pattern was similar whether using the observed serial interval distribution from the outbreak data or using a parametric probability distribution.

Our study has some limitations. Although data collection were standardised it is often difficult to assess the accuracy of the date and place that patients were when they became ill. Specifically, accurate information on patients' positions on a ward was available for 44 percent of the outbreaks. The spatial analysis was undertaken on 65 outbreaks. In addition to the sensitivity analysis we also analysed the data by including outbreaks where onset dates of illness were complete but data on patient location were incomplete (where fewer than 10% of patient data on position was incomplete, 85 outbreaks). We dealt with missing values by allocating a completely separate bay for patients with missing data on location at time of onset. This approach is conservative in that it would underestimate the impact of proximity. Secondly, we removed the patients from the outbreaks if positional information was missing. Despite this limitation, the additional models indicated that that the results are robust to different assumptions about missing data which is evidenced by slightly higher probabilities obtained when using records with complete information only (Table 1 and figure_1_supplement). As a check to demonstrate that the results were not an artefact of the statistical methods, we also ran the models on data where patient position was randomly assigned. This showed no pattern and the proximity measures were not significant for any of these models. In our analysis the estimation of P_k depends upon outbreak size. However we are not interested in the absolute values of P , only in how the value of P calculated with real proximity data compares with the value calculated with randomly generated proximity data (based on a permutation of the bay identities) which will be affected in the same way by outbreak sizes. We also performed a sensitivity analysis, normalising P_k by dividing it by the number of branches in the transmission tree for each network. This gives equal weight to

each outbreak and allows P_k to be interpreted as the probability that two linked cases were in the same bay. This did not change the results of the analysis; the P metric still fell well outside of the measure one would expect from the random simulations ($p = 0.004$, data not shown).

We used more than one approach to modelling the infection trees because of the lack of data on serial interval in norovirus outbreaks. Heijne *et al*'s method used data from child siblings at home. This was a useful starting point but is unlikely to be applicable to transmission in a hospital setting. Therefore, we derived gamma distributions for serial intervals from half a day to four days. The average incubation period for norovirus is considered to be between 24 hours to 48 hours [4, 18]. In our analysis the serial intervals of up to 2.5 days is likely to be a more appropriate time period in a hospital setting, than the analysis from Heijne *et al*.

Molecular analysis of stool samples could more definitively link outbreaks, which can help to reveal transmission networks [19, 20]. For example, in this study we have assumed that each ward outbreak was distinct, i.e. all cases within a ward were part of a chain of transmission, but this may not necessarily be true. It is possible for multiple introductions to occur, and some outbreaks may have spread from one ward to another. Genetic characterisation of samples from each ward during possible multiple outbreaks of norovirus would shed light on transmission events and lead to further insight about the direction of transmission, including the possibility that the virus can be moved around the hospital.

Our study focused on patients rather than staff. Our hypothesis was that symptomatic patients who vomit are most likely to contaminate the area close to them and other patients in their vicinity. Obtaining data on staff movements is much more complicated and would only really be practical in a detailed prospective study.

The importance of spatial proximity in propagating transmission is consistent with other recent studies[15, 21]. One study which used similar methods to calculate the infection trees

[15] suggests that symptomatic individuals are likely to be the drivers of outbreaks of norovirus in hospital settings. Furthermore, the effective reproductive number was significantly higher for symptomatic patients compared to that for symptomatic staff. Norovirus transmission between people in close contact during sport, both within and between teams, has also been shown to occur [22] as well as airborne transmission through explosive vomiting [23]. One study demonstrated that successive staff working on an aircraft in which a member of the public had vomited also became sick [11]. Norovirus has a low infectious dose [4, 24,25] shedding virus occurs during episodes of vomiting, where the virus can become aerosolized and expose others in the vicinity. Therefore, closing the bay quickly, preventing movement to and from that bay, and immediately paying attention to cleaning areas nearby to initial vomiting events are likely to be effective in preventing further spread. The index of suspicion for patients who become ill should be high and implementing infection control interventions should not be delayed until the results of sampling are received, because this would increase morbidity and prolong the outbreak. New guidelines on controlling outbreaks of norovirus in hospitals and care homes recently released in the U.K. [26] move away from the need to close wards and operate on a “manage within bays” principle. Our study has shown that patients in close proximity to symptomatic patients are at increased risk of becoming infected by these patients.

Conclusions

We have shown a clear role of spatial proximity in the transmission of norovirus in hospital outbreaks. Increasing barriers to movement between bays by closing affected bays promptly would be effective in preventing further spread.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Acknowledgements

We would like to thank the infection control staff at the NHS hospitals for providing the data used in this study.

Funding statement

No financial support was received for this study. All authors report no conflicts of interest relevant to this article

For peer review only

References

1. Patel MM, Hall AJ, Vinje J, et al. Noroviruses: a comprehensive review. *J Clin Virol* 2009;**44**:1-8
2. Phillips G, Tam CC, Conti S, et al. Community incidence of norovirus-associated infectious intestinal disease in England: improved estimates using viral load for norovirus diagnosis. *Am J Epidemiol*. 2010;**171**:1014-22
3. Tam CC, Rodrigues LC, Viviani L, et al. Longitudinal study of infectious intestinal disease in the UK (IID2 study): incidence in the community and presenting to general practice. *Gut* 2012;**61**:69-77
4. Estes MK, Prasad BV, Atmar RL. Noroviruses everywhere: has something changed? *Curr Opin Infect Dis*. 2006;**19**:467-74
5. Harris JP, Lopman BA, O'Brien SJ. Infection control measures for norovirus: a systematic review of outbreaks in semi-enclosed settings. *J Hosp Infect*. 2010;**74**:1-9
6. Lopman BA, Reacher MH, Vipond IB, et al. Epidemiology and cost of nosocomial gastroenteritis, Avon, England, 2002-2003. *Emerg Infect Dis*. 2004;**10**:1827-34
7. Danial J, Cepeda JA, Cameron F, et al, Templeton KE. Epidemiology and costs associated with norovirus outbreaks in NHS Lothian, Scotland 2007-2009. *J Hosp Infect*. 2011;**79**:354-8
8. Marks PJ, Vipond IB, Carlisle D, et al, Evidence for airborne transmission of Norwalk-like virus (NLV) in a hotel restaurant. *Epidemiol Infect*. 2000;**124**:481-7
9. Marks PJ, Vipond IB, Regan FM, et al, A school outbreak of Norwalk-like virus: evidence for airborne transmission. *Epidemiol Infect*. 2003;**131**:727-36
10. O'Neill PD, Marks PJ. Bayesian model choice and infection route modelling in an outbreak of Norovirus. *Stat.Med*. 2005;**24**:2011-24
11. Thornley CN, Emslie NA, Sprott TW, et al, Recurring norovirus transmission on an airplane. *Clin Infect Dis*. 2011;**53**:515-20

12. Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *Am J Epidemiol.* 2004;**160**:509-16

13. Heijne JC, Teunis P, Morroy G, et al. Enhanced hygiene measures and norovirus transmission during an outbreak. *Emerg Infect Dis.* 2009;**15**:24-30

14. Heijne JC, Rondy M, Verhoef L, et al. Quantifying transmission of norovirus during an outbreak. *Epidemiology* 2012;**23**:277-84

15. Sukhrie FH, Teunis P, Vennema H, et al. Nosocomial Transmission of Norovirus Is Mainly Caused by Symptomatic Cases. *Clin Infect Dis.* 2012;**54**:931-7.

16. R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.

17. Kenah E, Lipsitch M, Robins JM. Generation interval contraction and epidemic data analysis. *Math Biosci.* 2008;**213**:71-9

18. Patel MM, Hall AJ, Vinje J, et al. Noroviruses: a comprehensive review. *J Clin Virol.* 2009;**44**:1-8

19. Allen DJ, Gray JJ, Gallimore CI, et al. Analysis of amino acid variation in the P2 domain of the GII-4 norovirus VP1 protein reveals putative variant-specific epitopes. *PLoS ONE* 2008;**3**:e1485 doi:10.1371/journal.pone.0001485

20. Sukhrie FH, Beersma MF, Wong A et al, Using molecular epidemiology to trace transmission of nosocomial norovirus infection. *J Clin Microbiol* 2011;**49**:602-6

21 Teunis, P, Heijne, JCM, Sukhrie F, et al, Infectious disease transmission as a forensic problem: who infected whom? *J R Soc Interface* 2013;**6**:10 20120955. doi: 10.1098/rsif.2012.0955. published first online 6 February 2013.

22. Desai R, Yen C, Wikswo M, et al. Transmission of norovirus among NBA players and staff, winter 2010-2011. *Clin Infect Dis* 2011;**53**:1115-7

- 1
2
3 23. Sawyer LA, Murphy JJ, Kaplan JE, et al. 25- to 30-nm virus particle associated with a
4 hospital outbreak of acute gastroenteritis with evidence for airborne transmission. Am J
5 Epidemiol 1988;**127**:1261-71
6
7
8
9 24. Teunis PF, Moe CL, Liu P, et al. Norwalk virus: how infectious is it? J Med Virol
10 2008;80:1468-76
11
12
13 25. Atmar RL, Estes MK, The epidemiologic and clinical importance of norovirus infection.
14 Gastroenterol Clin North Am 2006;**35**:275-90, viii
15
16
17 26. Guidelines for the management of norovirus outbreaks in acute and community health
18 and social care settings. 2012
19
20
21
22
23 http://www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1317131639453 accessed 15/04/2013
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure legends.

Figure 1. Serial interval distribution derived from onset dates of illness from observed outbreaks.

Figure 2. observed (dashed line) and distribution of expected (bars) proximity metrics for each serial interval.

For peer review only

Appendix 1 supplementary information

Table 1 Modelling approaches used in the analyses.

| |
|--|
| Serial interval derived from observed data on onset dates from hospital patients during outbreaks (N=149) |
| Analysis 1: (main analysis in results) Include only outbreaks with complete data on patient position at time of onset. (N=65) |
| Sensitivity analyses |
| Analysis 2 Include outbreaks with missing values on patient position at time of onset. Patients allocated to distinct bays (N=85) |
| Analysis 3 Include outbreaks with missing values on patient position at time of onset, but exclude observations with missing values on patient position, i.e. patients deleted from dataset (N=85) |
| Analysis 4 Remove all information on patient position and randomly allocate bays |
| Using gamma distributions of serial intervals from Heijne et al (From healthy population) |
| |

Figure 1_supplement. Comparison of results from modelling approaches, main analysis and sensitivity analyses.

R. Code.

This first part of the program creates functions that will be called in the main part of the program later on.

create the distance matrix function which codes 1 for patients sharing a bay and 0 for those in different bays

```
makeDistancematrix<-function(pid,bay){
  n<-length(pid)
  Dist<-matrix(data=0, nrow=n,ncol=n)
  for(i in 1:n){
    for(j in 1:n){
      if(i!=j & bay[i]==bay[j]) Dist[i,j]<-1
    }
  }
  return(Dist)
}
```

Create the simulated outbreak function this simulates one outbreak

```
sim1outbreak<-function(pk1,onset.times.percase){
  n<-dim(pk1)[1]
  number.of.possible.index.cases<-sum(onset.times.percase==min(onset.times.percase))
  indexcase<-1+as.integer((number.of.possible.index.cases)*runif(1))
  whoinfectswhom<-rep(NA,n)
  for(i in 1:n){
    probnosource=max(1-sum(pk1[i,]),0)
    source<-which(rmultinom(1,1,c(pk1[i,],probnosource)) ==1)
    if(source==n+1) source<-0
    whoinfectswhom[i]<-source
    if(i==indexcase) whoinfectswhom[i]<-0 # i.e. infected outside
  }
  return(whoinfectswhom)
}
```

Create the function to randomise the bays in which patients are located

```
permutebays<-function(bays){
  n<-length(bays)
  return(bays[order(runif(n))])
}
```

Create the function to calculate the infection trees

```
calcOutbreakDist<-function(whoinfectswhom,Dist){
  obDist=0
  for(i in 1:length(whoinfectswhom)){
    if(whoinfectswhom[i]!=0){
      obDist=obDist+Dist[i,whoinfectswhom[i]]
    }
  }
  return(obDist)
}
```

This section is the main program, it begins by reading in the data "data.csv" which is the comma delimited data file containing the data.

```
dataset<-read.table("data.csv",header=T,sep=",")
```

```

1
2
3 #to ensure that the dates in the csv file are interpreted as dates by R
4 dataset$dateonset<- as.Date(as.character(dataset$dateonset),"%d/%m/%Y")
5 S<-1000 # note this sets the number of simulated distance matrices to create
6 numobks<-length(unique(dataset$obnumb))
7 realcumulativeDist<-rep(0,numobks) # to calculate cumulative distance over all simulated outbreak
8 trees
9 permutedDistances<-matrix(data = rep(0,S*numobks), nrow = numobks, ncol = S)
10 for(obnumb in 1:numobks){
11   outbreak.id<-obnumb
12   onset.times.percase<- dataset$dateonset[which(dataset$obnumb==outbreak.id)]
13   unique.onset.times.percase<-unique(onset.times.percase)
14   unique.onset.times.percase<-1+unique.onset.times.percase - min(unique.onset.times.percase)
15   N<-length(onset.times.percase)
16   maxtimediff<-max(onset.times.percase)-min(onset.times.percase)
17   w<-rep(NA,maxtimediff+1) #probability that the serial is i-1 days
18   w[1]<-0.6377
19   w[2]<-0.1678
20   w[3]<-0.1074
21   w[4]<-0.0403
22   w[5]<-0.0000
23   w[6]<-0.0134
24   w[7]<-0.0000
25   w[8]<-0.0201
26   w[9]<-0.0000
27   w[10]<-0.0067
28   w[11]<-0.0000
29   w[12]<-0.0000
30   w[13]<-0.0000
31   w[14]<-0.0000
32   w[15]<-0.0000
33   w[16]<-0.0000
34   w[17]<-0.0000
35   w[18]<-0.0067
36   w[19]<-0.0000
37   w[20]<-0.0000
38   w[21]<-0.0000
39   w[22]<-0.0000
40   w[23]<-0.0000
41   w[24]<-0.0000
42   w[25]<-0.0000
43   w[26]<-0.0000
44   w[27]<-0.0000
45   w[28]<-0.0000
46   w[29]<-0.0000
47   w[30]<-0.0000
48   wij<-matrix(rep(0,N^2),nrow=N)
49   number.of.possible.index.cases<-sum(onset.times.percase==min(onset.times.percase))
50   for(i in 1:N){
51     j<-1
52     onsettime<-onset.times.percase[j]
53     while(onsettime<=onset.times.percase[i] && j<=N){
54
55       timediff<-onset.times.percase[i]-onset.times.percase[j]
56       if(i!=j){
57         if(i> number.of.possible.index.cases){

```

```

1
2
3     wij[i,j]<- w[timediff+1]
4   } else {
5     wij[i,j]<- w[timediff+1]*(number.of.possible.index.cases-1)/ (number.of.possible.index.cases)
6   }
7 }
8 j<-j+1
9 onsettime<-onset.times.percase[j]
10 }
11 }
12
13 # denoms is the sum of likelihoods that person i was infected by person i-1,i-2,...,1
14 denoms<-c(NULL,rep(0,N-1))
15 for(i in 1:N){
16   denomfori<-0
17   for(j in 1:N) {
18     denomfori<-denomfori+wij[i,j]
19   }
20   denoms[i]<-denomfori
21 }
22
23 pkl<-matrix(rep(0,N^2),nrow=N) #relative likelihood that k was infected by l
24
25 for(l in 1:N){
26   # print(l)
27   for(k in 1:N){
28     numerator<-wij[k,l]
29     denom<-denoms[k]
30     if(denom>0){
31       pkl[k,l]<-numerator/denom
32     } else {
33       pkl[k,l]<-0
34     }
35   }
36 }
37
38 for(i in 1:number.of.possible.index.cases) pkl[i,]<-pkl[i,]*(number.of.possible.index.cases-
39 1)/number.of.possible.index.cases
40
41 pid <- dataset$caseid[which(dataset$obnumb==outbreak.id)] #This limits the variable pid to one
42 outbreak
43
44 bay <- dataset$bay[ which(dataset$obnumb==outbreak.id)] #This limits the variable bay to one
45 outbreak
46
47
48 Dist <- makeDistancematrix(pid,bay)
49   for (i in 1:S) {
50     oneoutbreaksim<-sim1outbreak(pkl,onset.times.percase) #a possible reconstruction of outbreak -
51 who infects whom.
52     realDist<-calcOutbreakDist(oneoutbreaksim, Dist)
53     realcumulativeDist[obnumb]<-realcumulativeDist[obnumb] + realDist
54     permD<-makeDistancematrix(pid,permutebays(bay)) #this is for the permuted bays
55     DistwithpermutedDistancematrix<-calcOutbreakDist(oneoutbreaksim, permD)
56     permutedDistances[obnumb,i]<-permutedDistances[obnumb,i]+ DistwithpermutedDistancematrix
57   } # end for i in 1:S
58
59
60

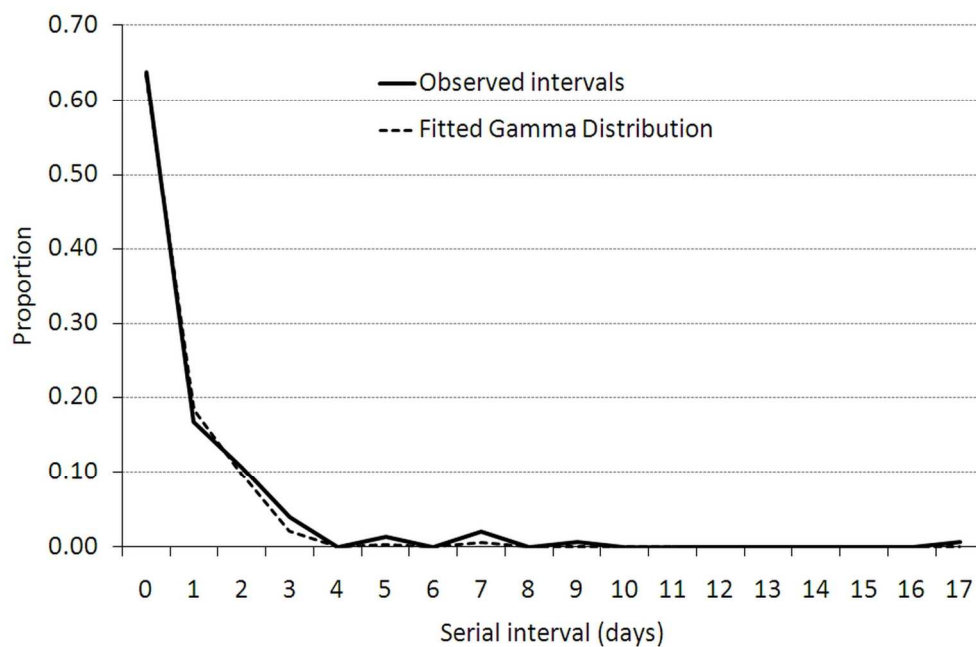
```

```
1
2
3   realcumulativeDist[obnumb]<-realcumulativeDist[obnumb]/S
4   permutedDistancesAllOutbreaks<-colSums(permutedDistances)
5   test.statistic<-sum(realcumulativeDist) # for all outbreaks
6   pvalue.2sided<-
7   2*min(sum(permutedDistancesAllOutbreaks<=test.statistic)/length(permutedDistancesAllOutbreaks),
8   sum(permutedDistancesAllOutbreaks>=test.statistic)/length(permutedDistancesAllOutbreaks))
9
10  } #end for(obnumb in 1:numobks)
11
12  -----
13  # Note this section can be used instead of the observed serial interval distribution if using the gamma
14  distribution estimate of serial intervals
15
16  w[1]<-pgamma(0.5,shape=shape,scale=scale) # probability of a serial interval recorded as 0 days (i.e.
17  0 to .5 days)
18  for(i in 1:maxtimediff) w[i+1]<-pgamma(i+0.5,shape=shape,scale=scale)-sum(w[1:i])
19  -----
```

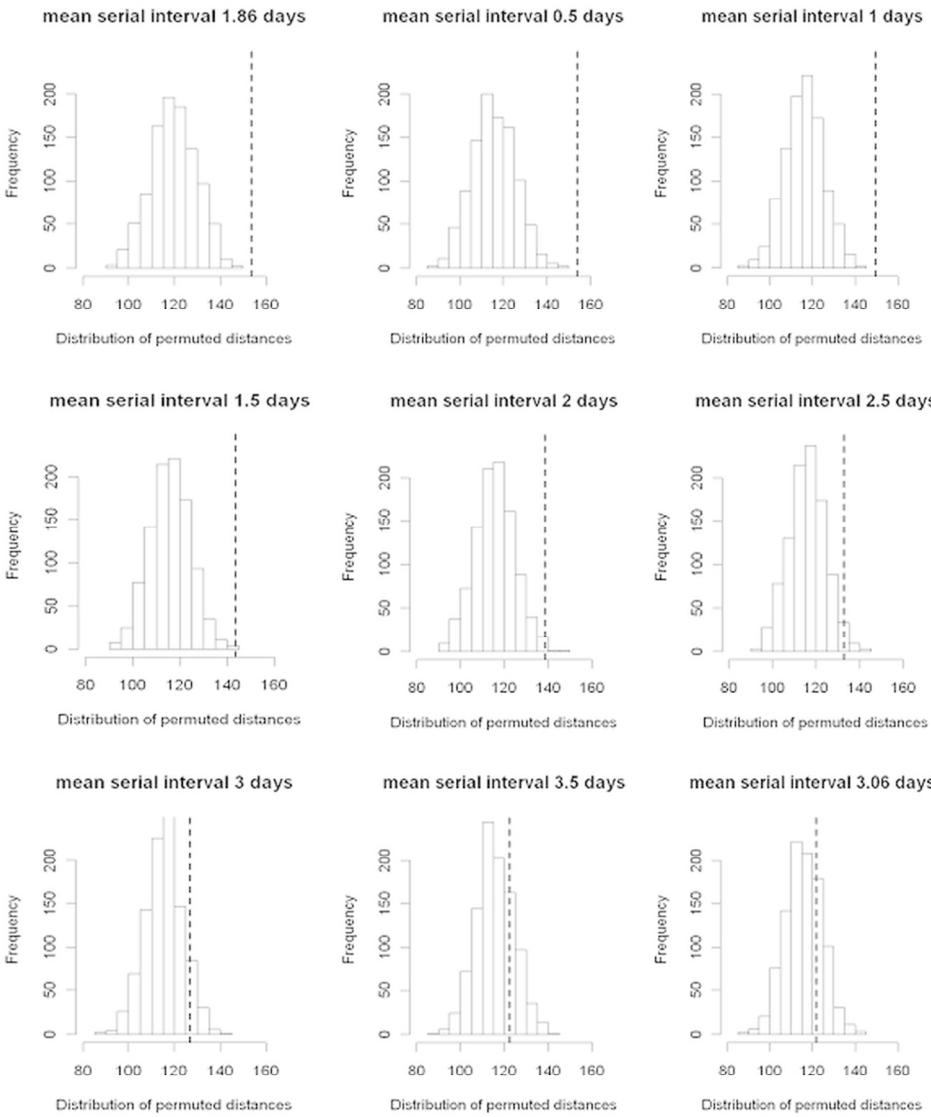
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Sample data.

| obnumb | dateonset | Bed number | Bed number | bay | caseid |
|--------|------------|------------|------------|-----|--------|
| 1 | 22/12/2007 | 7 | 7 | 1 | 1 |
| 1 | 25/12/2007 | 5 | 5 | 3 | 2 |
| 1 | 25/12/2007 | 3 | 3 | 6 | 3 |
| 1 | 25/12/2007 | 6 | 6 | 1 | 4 |
| 1 | 26/12/2007 | 6 | 6 | 3 | 5 |
| 1 | 26/12/2007 | 3 | 3 | 1 | 6 |
| 1 | 27/12/2007 | 7 | 7 | 3 | 7 |
| 1 | 28/12/2007 | 4 | 4 | 1 | 8 |



serial interval distribution



observed (dashed line) and distribution of expected (bars) proximity metrics for each serial interval.

Does spatial proximity drive norovirus transmission during outbreaks in hospitals?

Article focus:

- Published literature on norovirus outbreaks does not provide clear evidence of the effectiveness of infection control measures
- Improved understanding of how norovirus spreads in closed environments could lead to better infection control procedures.
- This study uses statistical modelling methods to assess whether patients in close proximity are at increased risk of contracting norovirus during outbreaks in hospitals

Key Messages:

- We have shown a clear role of spatial proximity in the transmission of norovirus in hospital outbreaks.
- Patients who are in the same bay as patients who become ill have a higher probability of becoming ill compared to patients in a different bay.
- Increasing barriers to movement between bays by closing affected bays promptly would be effective in preventing further spread

Strengths and Limitations of this study:

- Provides an estimation of serial interval, and assessment of significance of patient proximity in spreading norovirus within hospitals.
- Different modelling approaches showed consistent results.
- A weakness is that although data collection were standardised it is often difficult to assess the accuracy of the information on patients' positions on a ward.

STROBE Statement—checklist of items that should be included in reports of observational studies

| | Item No | Recommendation |
|---------------------------|---------|--|
| Title and abstract | 1 | (a) Indicate the study’s design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found |
| Introduction | | |
| Background/rationale | 2 | Explain the scientific background and rationale for the investigation being reported |
| Objectives | 3 | State specific objectives, including any prespecified hypotheses |
| Methods | | |
| Study design | 4 | Present key elements of study design early in the paper |
| Setting | 5 | Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection |
| Participants | 6 | (a) Cohort study—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up Case-control study—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls Cross-sectional study—Give the eligibility criteria, and the sources and methods of selection of participants (b) Cohort study—For matched studies, give matching criteria and number of exposed and unexposed Case-control study—For matched studies, give matching criteria and the number of controls per case |
| Variables | 7 | Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable |
| Data sources/ measurement | 8* | For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group |
| Bias | 9 | Describe any efforts to address potential sources of bias |
| Study size | 10 | Explain how the study size was arrived at |
| Quantitative variables | 11 | Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why |
| Statistical methods | 12 | (a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) Cohort study—If applicable, explain how loss to follow-up was addressed Case-control study—If applicable, explain how matching of cases and controls was addressed Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses |

Continued on next page

Results

| | | |
|------------------|-----|---|
| Participants | 13* | (a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed (b) Give reasons for non-participation at each stage (c) Consider use of a flow diagram |
| Descriptive data | 14* | (a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders (b) Indicate number of participants with missing data for each variable of interest (c) <i>Cohort study</i> —Summarise follow-up time (eg, average and total amount) |
| Outcome data | 15* | <i>Cohort study</i> —Report numbers of outcome events or summary measures over time <i>Case-control study</i> —Report numbers in each exposure category, or summary measures of exposure <i>Cross-sectional study</i> —Report numbers of outcome events or summary measures |
| Main results | 16 | (a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period |
| Other analyses | 17 | Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses |

Discussion

| | | |
|------------------|----|--|
| Key results | 18 | Summarise key results with reference to study objectives |
| Limitations | 19 | Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias |
| Interpretation | 20 | Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence |
| Generalisability | 21 | Discuss the generalisability (external validity) of the study results |

Other information

| | | |
|---------|----|---|
| Funding | 22 | Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based |
|---------|----|---|

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at www.strobe-statement.org.